2016

The Hot Hand vs. Cold Hand on the PGA Tour

Arkes, Jeremy

Jeremy Arkes. "The Hot Hand vs. Cold Hand on the PGA Tour." International Journal of Sport Finance . May 2016, Vol. 11 Issue 2, p. 99-113.
http://hdl.handle.net/10945/59138

2016

The Hot Hand vs. Cold Hand on the PGA Tour

Arkes, Jeremy

Jeremy Arkes, "The Hot Hand vs. Cold Hand on the PGA Tour." International Journal of Sport Finance . May 2016, Vol. 11 Issue 2, p. 99-113.
http://hdl.handle.net/10945/59138

International Journal of Sport Finance, 2016, **11**, 99-113, © 2016 West Virginia University

The Hot Hand vs. Cold Hand on the PGA Tour

Jeremy Arkes[1]

[1] Naval Postgraduate School

Jeremy Arkes, PhD, is an associate professor of economics in the Graduate School of Business and Public Policy at the Naval Postgraduate School. His research interests are military manpower, sport economics, behavioral economics, substance use, and divorce effects on children.

Abstract

This paper examines tests for hot- and cold-hand effects in men's professional golf, based on score relative to par over sets of 3, 6, 9, and 18 holes. Using controls for each player's annual performance and the difficulty of a set of holes on a given day, I find no evidence for a hot-hand effect. In contrast, I find evidence supporting the existence of a cold-hand effect—that is, a poor performance on one set of holes leads to a worse performance on the next set of holes. Simulations demonstrate that the cold-hand effect is quite large.

Keywords: hot hand, cold hand, golf

Introduction

There has been much research devoted to the "hot hand" in sports. The hot hand is a period of elevated performance for an individual player. The most studied sport for the hot hand—and the one that has had the greatest recent transformation in the results—has been basketball. Starting with Gilovich et al. (1985), a series of articles over 25 years had found no evidence for the hot hand in basketball and claimed that the hot hand was a myth. That is, when players or fans believe a player has the hot hand, the argument goes, they are just misperceiving natural statistical variation as having patterns.

In the last five years, however, new research has cast doubt on this conclusion. Arkes (2010) developed a player-fixed-effects model that allowed all players to be included in one model (in contrast with prior studies that examined one player at a time), thus allowing for a much larger sample. Using free throw data, he found that making the first of two free throws leads to a 3-percentage-point increase in the probability of making the second free throw. This is not a large effect, but it was the first evidence for the hot hand in basketball. Recent studies using similar methods also found evidence for hot-hand effects in basketball (Bocskocsky et al., 2014; Miller & Sanjuro, 2014). Further studies by Stone (2012) and Arkes (2013) demonstrated that the methods used in prior studies—which includes, by extrapolation, the more recent articles with evidence for the hot hand—were subject to a downward bias, largely from measurement error. The prior studies, they found, would have had a low probability of detecting the hot hand (i.e., statistical significance) even if there were a strong and relatively frequent hot hand effect.

The main problem with the studies, Stone (2012) argues, is equivalent to measurement error. That is, the typical measure for whether a player is hot is whether the player hit the prior shot. But, the hot hand means that a player is shooting at an elevated level, not hitting every shot. Thus, due to natural variation, a player could be hot and miss the shot, and a person could be in a normal state and make the shot. Measurement error would then bias the estimated hot-hand effect towards zero. This problem with the basketball hot-hand studies would apply to models on the hot hand for any sport, given the large role of randomness in sports outcomes and the inevitable measurement error. And, there does not appear to be any solution to this inherent bias other than increase the sample size to generate enough power to detect a hot-hand effect, if one indeed were to exist.

Further problems in the hot-hand studies were recently discovered by Miller and Sanjurjo (2015). They argued that there is selection bias in the studies in that the common practice of comparing performance after streaks of made shots to performance after streaks of missed shots takes out of the sample part of the hot-hand period. They then corrected for the selection bias with data from the initial hot-hand study (Gilovich, 1985) and find a fairly large and significant hot-hand effect.

One other problem with some of the literature on the basketball hot hand is the possibility of endogenous responses. That is, if a player is hot, the defense may adjust by various methods, such as shifting the best defender on the hot player or double-teaming the player. This would make it more difficult to detect the hot hand.

Addressing the issue of endogenous responses, some studies have studied the hot hand in sports that have no defense, such as horseshoes (Smith, 2003) and bowling (Dorsey-Palmateer & Smith, 2004), although these studies had weak power. Baseball is a sport with some defense, but limited endogenous responses of just pitching around a hitter. Most of the research has found no evidence for a hot hand in baseball (Albert & Bennett, 2003; Albright, 1993; Vergin, 2000); but, a recent study with two million observations found strong evidence for the hot hand using a variety of offensive measures (Zwiebell & Green, 2014).

Golf is another sport with no possibility of an endogenous response. In fact, golf could be an ideal sport to examine the hot hand. There is a standard scoring and shot system, so virtually everyone has the same number of holes and course difficulty for measuring performance. And with this standard system of 18 outcomes (holes) per round, performance over several holes can be combined to reduce the impact of randomness and measurement error. That is, combining holes would provide a more accurate measure of the level of performance for a player, which reduces the role of randomness, thereby reducing (but by no means eliminating) the measurement error. That said, out of roughly four shots per hole, one bad shot could lead to a bad score on the hole despite two or three other excellent shot. Thus, a hot hand in golf would likely require a hot hand in perhaps a few distinct shot types: drives, approaches, and putts. In contrast, a cold hand could result from not doing well in just one type of shot, say putting.

The drawback to using golf is that it is more removed from what we think of as the hot hand in terms of adrenaline and performing based on instinct and natural split-second reactions. Golf shots involve time to think about how to approach a shot and allow for practice swings. On the other hand, one attribute that could contribute to the hot hand—confidence—could very well determine success in golf. Nevertheless, it comes down to an empirical question of whether a player performs better (worse) if he had performed well (poorly) in the prior round or prior set of holes.

In this study, I apply a similar fixed-effects model as in Arkes (2010) to examine whether there is a hot hand and a cold hand in golf—particularly on the Professional Golf Association (PGA) tour. A recent event highlights the possibility of the hot hand in golf. At the 2014 Travelers Championship (in Connecticut), Kevin Streelman was 2-over-par through the first seven holes of the final round. He proceeded to one-putt each hole on the back nine, getting birdies on the last seven holes. With this performance, he won the tournament by one stroke. Of course, it is possible that this was natural variation, and Streelman was just lucky.

In the first studies on the hot hand in golf, Clark (2003a, 2003b, 2005) estimated Wald-Wolfowitz runs tests for round-to-round performance. He defined a strong performance as having a "par or better" round. He found some evidence for correlated performances for individual players (meaning more runs of consecutive strong performances or consecutive non-strong performances), but he attributed this to differences in course difficulty. Thus, he concludes that there is no evidence for the hot hand in golf. Connolly and Rendleman (2008) also examined round-to-round hot-hand effects as part of a larger study decomposing performance. They found that 9% of 253 golfers examined over the 1998-2001 period exhibited positive autocorrelation (indicating a hot hand), although that is not much more than would be expected by chance. Rosenqvist and Skans (2014) used regression discontinuities and found evidence for confidence effects from making a tournament cut one week on performance in the next tournament (a week later). While the time between tournaments may be too long to be considered a typical "hot hand," the results are consistent with a hot-hand effect.

The most relevant article to this research is Livingston (2012). He examined separate effects of bad (above par) and good (below par) performance on one hole on the probability of a bad or good performance on the next hole. He limited the sample to one tournament each from 2006 for the PGA Tour, the LPGA Tour, the Champions (age 50 and older) Tour, and the Nationwide Tour (a "minor league"), which provides samples of between 4,000 and 9,000 observations. And, he controlled for the average score on the hole and the player's average score for the 2006 season. The measures of hot or cold play were four variables based on streaks of one good, two good, one bad, or two bad holes. He only found consistent evidence for a hot or cold hand in the Nationwide Tour. Most notably, he found no evidence for any hot- or cold-hand effect for the PGA Tour, which is the tour I focus on.

In this article, I build upon the prior research in several ways. First, rather than examine one player at a time—as in all articles other than Livingston (2012)—I apply Arkes' (2010) player-fixed-effects model, which should give the model much greater power from the larger number of observations. This model can also hold constant difficulty for each course or set of holes on a given day. Second, I use a measure of performance that has greater variation than what prior studies used: I use score relative to par (over 3, 6, or 9 holes or a whole round). Thus, there are far greater possible values for the level of performance compared to just the four possible values used in

Livingston (2012). The greater variation in the prior level of performance offers further power for detecting any hot- and cold-hand effects. Third, understanding the inevitable measurement-error bias, I use simulations with various levels of the hot hand and cold hand that attempt to gauge what the estimated effects in this study could actually represent.

I find no evidence for the hot hand, but I do find strong evidence for a cold hand, as measured by score relative to par from one set of 3, 6, or 9 holes to the next within the same round and from one set of 18 holes to the next within the tournament. The estimates on the cold hand are consistent with having very large reductions in performance, equivalent to going from around the third-highest decile for average performance in a tournament to the bottom decile.

Methods

Data

The data come from the men's PGA Tour Headquarter's Shotlink data. The data provide information on every shot from each PGA event. The information includes the tournament, the course, the round, the hole, the player, the par score for each hole, the player's score on each hole, the distance from the hole for each shot, whether each shot was made, and many other factors. The data are available from the 2003 season to the current season. However, with 2014 being just partially completed at the time of analysis, I just use data over the 11-year period of 2003–2013.

Excluding one tournament with a special format (the Bob Hope Classic), there are 3,280,468 holes completed by players in official PGA tour events in this period, for 1,757 different players. I limit the sample to completed rounds for the top 200 golfers for each of the 11 years. This produces 2,779,452 holes played, or 154,414 rounds of golf. I only included data from full rounds of golf, so rounds of golf cut short due to injury or sudden-death play-offs would be excluded.

Samples

I conduct the analysis for various sets of holes: 3, 6, 9, and 18. Sets of 18 holes (or full rounds of golf) are analyzed within the same tournament, with the idea that it is less likely that the hot or cold hand, if it were to exist, would last from one tournament to the next. The analysis of sets of three, six, or nine holes would only be for the hot/cold hand within a single round of golf. Thus, for the nine-hole analysis, the model would examine how performance on the first nine holes translates into performance on the second nine holes. Thus, performance on the first nine holes in a given round would not be included as a dependent variable, but rather just an explanatory variable for the outcome of performance on the second nine holes of a round. The same restrictions apply for the six-hole and three-hole analysis, as the first set of holes in a round is excluded. I do not examine the hot/cold hand from one hole to the next because there could be prospect-theory effects in that performance on a given hole could cause a person to change effort or risk to make up for poor performance or to try to preserve the gains from a strong performance (Stone & Arkes, 2016).

Table 1 shows the descriptive statistics for each sample, based on either 3, 6, 9, or 18 holes. The sample size of 772,070 for three-hole performance is based on 926,484 sets of three holes in the 2,779,452 total holes played in the sample and eliminating one-sixth of those (the first three holes in a given round). The number of observations for the six-hole analysis is based on the same type of calculation. The 154,414 observations for the nine-hole analysis is the number of rounds played. The 18-hole analysis is then smaller due to the first round of any tournament not being used as a dependent variable. The average score over 18 holes is 0.274 strokes below par. Of the full rounds of 18 holes, 49% were below par and 38% were above par. Not surprisingly, as the set of holes analyzed goes from three holes to 18 holes, the percentage of rounds below par and above par increase.

About 38% of rounds start off on the back nine (with hole 10) rather than the front nine (with hole 1). Adjustments are made to make sure the performance is measured on various sets of holes sequentially.

Empirical Model

The most common method to test for the hot hand in sports is to analyze one player at a time, which most studies have done. This analysis uses player fixed effects, as introduced in Arkes (2010), to incorporate all players in one model. But, in the case of golf, more controls are needed. If a positive correlation between performance over consecutive sets of holes or rounds were found, it could due to the difficulty of the course (or sets of holes) or weather conditions for a particular day. This would involve a vast number of fixed effects, so the model just includes the average score relative to par for the set of (3, 6, 9, or 18) holes being examined—that is, the set of holes that an individual player's score is representing is the dependent variable. The model is the following:

$$\left(S_{\mathrm{i,y,t,c,r,h}}\right) = \beta(S_{\mathrm{i,y,t,c,r,h-1}}) + \mu_{i,y} + \gamma\bar{S}_{\mathrm{y,t,c,r,h}} + \varepsilon_{\mathrm{i,y,t,c,r,h}} \quad (1)$$

where *i* refers to the player, *y* refers to the year, *t* refers to the tournament, *c* refers to the course, *r* refers to the round, *h* refers to the set of holes, $S_{i,y,t,c,r,h}$ represents the player's score relative to par for the set of holes being examined, $S_{i,y,t,c,r,h-1}$ is the player's score relative to par for the prior set of holes (this is a lagged dependent variable), $\mu_{i,y}$ is the player-year fixed effects, $\bar{S}_{y,t,c,r,h}$ is the average score for year-tournament-course-round-hole (which would be for 3, 6, 9, or 18 holes on a given day), and $\varepsilon_{i,y,t,c,r,h}$ is the the error term.

The key parameter is β, which indicates how much a one-stroke improvement in performance for one set of holes translates into any change in performance in the next

Table 1. Summary Statistics

	3 holes (*n*=772,070)	6 holes (*n*=308,838)	9 holes (*n*=154,414)	18 holes (*n*=105,617)
Score relative to par	-0.030 (1.189)	-0.057 (1.716)	-0.079 (2.140)	-0.274 (3.174)
Whether player scored below par for set of holes	0.346 (0.476)	0.411 (0.492)	0.438 (0.496)	0.491 (0.500)
Whether player scored above par for set of holes	0.293 (0.455)	0.349 (0.477)	0.370 (0.483)	0.383 (0.486)

Note: The standard deviation is in parentheses.

set of holes. Standardization for the number of holes being evaluated (e.g., performance over 3, 6, 9, or 18 holes) is not necessary because both the dependent variable and the lagged dependent variable are based on the same number of holes.

Another specification uses two variables instead of the player's score on the prior set of holes: (1) (score below par)*(whether the player was below par); and (2) (score above par)*(whether the player was above par). The excluded category that has zeroes for these two variables would be for those with a total score of par for the set of holes. Using these two variables can distinguish between the hot hand (for scores below par) and the cold hand (for scores above par).

The player-year fixed effects are included to control for the average performance of a player in a given year. The average score for the year-tournament-course-round for each set of holes are meant to control for the difficulty of the set of holes being examined for a given round, which should do well in controlling for the effects of weather conditions. When the difficulty of the sets of holes varies over the course of a day (or over a few days if weather causes a round to occur over multiple days), the model will not be able to control for that.

With all of the controls, the interpretation of the estimates is that it represents whether players do better (or worse) on one set of holes if they had done better (or worse) on the prior set of holes, holding constant the average performance on the set of holes by all players on that set of holes that day, and how well the player did over the given year.

It is possible that some players are subject to the hot or cold hand, while others are not. With fixed effects, the estimated hot-hand and cold-hand effects are naturally estimated as the weighted average of these effects across players, with the weights being based on the number of observations for each player and the variation in his performance on a given set of holes (Gibbons & Suarez Serrato, 2011). Because variation in performance is not very different across players, the natural weight each player has in the determination of the overall coefficient estimate is mostly driven by the number of observations for a given player.

How should the estimated hot- and cold- hand effects vary based on the number of holes being evaluated? There are two opposing effects. The fleeting nature of the hot hand in sports would cause a larger set of holes being examined to result in a smaller hot-hand effect. However, the role of randomness (and thus measurement error in misidentifying when a player was hot) would be smaller with a larger set of holes, resulting in a larger estimated hot-hand effect. The results will indicate which effect dominates as the set of holes becomes larger.

Results

Table 2 shows the results for the models testing for the hot and cold hand in golf. The samples are based on 3, 6, and 9 holes within a round and 18 holes within a tournament (testing for the round-to-round hot hand). For the 18-hole sample, the first round is excluded; and for the rest, the first set of holes (e.g., first nine holes for the nine-hole sample) in a round is excluded because the prior set of holes would have been from the prior round (most likely the previous day).

The odd columns use just the score relative to par, while the even columns separate scores below and above par. In the odd columns, the coefficient estimate on the score

Table 2. Estimates on the Score-Relative-to-Par for the Prior Set of Holes

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	3-hole performance		**6-hole performance**		**9-hole performance**		**18-hole performance**	
Performance on prior set of 3, 6, 9, or 18 holes								
Score relative to par	0.0093*** (0.0011)		0.0184*** (0.0017)		0.0257*** (0.0025)		0.0212*** (0.0030)	
Score below par		0.0018 (0.0020)		0.0049 (0.0032)		0.0071 (0.0046)		-0.0046 (0.0053)
Score above par		0.0161*** (0.0018)		0.0312*** (0.0031)		0.0434*** (0.0045)		0.0509*** (0.0058)
Average score for the given holes for the given round								
3 holes	3.044*** (0.009)	3.043*** (0.009)						
6 holes			6.074*** (0.027)	6.073*** (0.027)				
9 holes					9.235*** (0.059)	9.239*** (0.059)		
18 holes							18.56*** (0.12)	18.59*** (0.12)
# Obs.	772,070	772,070	308,828	308,828	154,414	154,414	105,617	105,617
R^2	0.131	0.131	0.164	0.164	0.175	0.175	0.248	0.248

Note: The dependent variable is the score relative to par on the set of 3, 6, 9, or 18 holes. The prior set of holes is the same number of holes as the dependent variable—i.e., prior round for analysis of full round, prior nine holes for the analysis on nine holes, etc. Standard errors are in parentheses. The model also includes player-year fixed effects and a constant term.
*** $p<0.01$, ** $p<0.05$, * $p<0.10$

relative to par on the prior set of holes is statistically significant in each of the models and increasing with larger sets of holes, up to nine holes, but it is lower for 18 holes.

The estimates suggest a hot-hand effect, but they could also represent a cold-hand effect. The models with separate variables for scores below par and above par (in the even columns) distinguish between the hot- and cold-hand effects. And these estimates show no significant evidence for the hot hand. That is, there is no evidence that success in one set of holes leads to a better-than-normal performance on the next set of holes. In contrast, there is evidence for a cold-hand effect. For each stroke above par on the prior set of 3, 6, 9, or 18 holes, players score an estimated 0.016, 0.031, 0.043, and 0.051 strokes above par on the current set of holes, with $p<0.001$ for all estimates.

Table 3. A Comparison of Estimates from Original Model and Excluding Rounds 2 and 4

	(1)	(2)	(3)	(4)	(5)	(6)
	3-hole performance		6-hole performance		9-hole performance	
VARIABLES	Original	Excluding Rds 2 & 4	Original	Excluding Rds 2 & 4	Original	Excluding Rds 2 & 4
Performance on prior set of 3, 6, or 9 holes						
Score below par	0.0018	-0.0009	0.0049	0.0049	0.0071	0.0057
	(0.0020)	(0.0028)	(0.0032)	(0.0045)	(0.0046)	(0.0065)
Score above par	0.0161***	0.0105***	0.0312***	0.0205***	0.0434***	0.0377***
	(0.0018)	(0.0026)	(0.0031)	(0.0043)	(0.0045)	(0.0063)
Average score for the given holes for the given round						
3 holes	3.043***	3.032***				
	(0.009)	(0.013)				
6 holes			6.073***	6.044***		
			(0.027)	(0.038)		
9 holes					9.239***	9.100***
					(0.059)	(0.083)
# Obs.	772,070	392,840	308,828	157,136	154,414	78,568
R^2	0.131	0.132	0.164	0.169	0.175	0.184

Note: The dependent variable is the score relative to par on the set of 3, 6, or 9 holes. The prior set of holes is the same number of holes as the dependent variable—e.g., prior nine holes for the analysis on nine holes. Standard errors are in parentheses.
*** $p<0.01$, ** $p<0.05$, * $p<0.10$

That is, performing poorly on one set of holes leads to worse performance on the next set of holes. These do not seem like large effects, but as shown in the next section, due to attenuation bias from measurement error, these small estimates are actually indicative of very large effects.

One concern may be that as the cut approaches at the end of the second round and as the final standings become clearer during the fourth (final round), players may change their strategy to become more cautious or more risky, depending on where they stand. Baldson (2013) finds evidence for changes in behavior, as being below the cutoff line for the cut (in the last five holes before the cut) was associated with a higher probability of getting both an above-par and below-par score. However, Baldson (2013) found no clear pattern of changes in risk behavior based on position in the last few holes of the tournament.

Nevertheless, to reduce the possibility that such changes in strategy affect the estimation of the hot and cold hand, I estimate the model excluding the second and fourth rounds. I can only do so for the three-hole, six-hole, and nine-hole analysis, as the 18-

hole analysis examines effects across rounds, so it would always include either the second or fourth round.

Table 3 shows a comparison of estimates from the new models that exclude data from rounds 2 and 4 and the original models in Table 2 that separate score below par and score above par. All of the estimated effects of score above par on the prior set of holes are smaller in the models that exclude rounds 2 and 4 (in the even columns) than the original estimate (in the odd columns). While they remain strongly significant, the differences in estimates suggest that changes in risky play as described by Baldson (2013)—that is, taking greater chances when behind—could explain up to about 35% of the cold-hand effect. The estimated effect of scores below par are slightly reduced for three- and nine -hole performance and do not change for six-hole performance. They all remain insignificant.

Simulation

The lack of evidence for a hot-hand effect and the relatively small estimates for the cold-hand effect may hide what could be a real hot-hand effect and a more pronounced cold-hand effect. As discussed earlier, Stone (2012) and Arkes (2013) demonstrate, with simulation data, that the existence of the hot hand would be difficult to detect and understated based on inherent problems with estimating hot-hand effects. In this section, I apply a method similar to Arkes (2013) to gauge what level of reduced and elevated performance would produce the cold-hand and hot-hand effects shown in Table 2. Furthermore, by having fixed hot-hand and cold-hand effects that have constant elevated and reduced levels of performance in a round, this simulation shows how the estimated hot/cold-hand effects and the probability of detecting them are lower when performance is measured over shorter periods of time due to measurement error. That is, the randomness in scores over shorter sets of holes creates more measurement error in that the performance on the holes is less representative of how well a player is playing than a longer set of holes. In other words, we cannot surmise that someone has the hot/cold hand based on performance on one set of holes. But elevated (or weak) performance over a wider set of holes would be stronger evidence for having the hot (or cold) hand.

This simulation is different from that in Arkes (2013) because performance on a hole is not a dichotomous outcome as it is in making a basketball shot. Rather, there are several possible outcomes, with the ones I model being (with the score relative to par in parentheses): eagle (-2), birdie (-1), par (0), bogey (+1), double bogey (+2), and triple bogey (+3). Double eagle and anything beyond triple bogey are rare enough that they can be assumed away, so there are six possible outcomes.

In basketball, having the hot hand would be associated with an elevated probability of making a shot. In golf, having the hot hand would be associated with a shift in the probability distribution for a given hole towards a lower score (e.g., less likely to make a bogey and more likely to make a birdie. Having the cold hand would have shifts in probabilities in the opposite directions).

Simulation Procedure

I perform separate simulations for the hot- vs. cold-hand effects. To characterize shifted probability distributions in a single variable, the simulation involves the following:

1. For all golfers who complete a tournament, meaning that they made the cut and finished 72 holes, I divide the players into deciles for each tournament over the 11 years of data, based on their scores.
2. I calculate the probability distribution for the six possible outcomes—from eagle (-2) to triple bogey (+3) for each decile.
3. I create an analysis sample based on the 2,779,452 holes (154,414 rounds) played. Players are all assigned a baseline distribution (some decile—e.g., 7th decile).
4. I then take a random sample of either 5% or 10% of all player-tournaments (even those where the player did not make the cut) and assign that player to be in the "hot hand" state for entire tournaments. In the cold-hand analysis, I assign 5% or 10% of player-tournaments to be in the "cold hand" state. I thus assign them to the highest-decile probability distribution (i.e., 10th decile) for the hot-hand simulation or the lowest-decile probability distribution (i.e., 1st decile) for the cold-hand simulation. By randomization, this means that roughly 5% or 10% of all rounds would be played with the hot hand or cold hand.
5. I then estimate the model

$$S_{i,h} = \beta S_{i,h-1} + \varepsilon_{i,h} \tag{2}$$

with the subscript *h* referring to the set of holes. Note that with everyone assigned the same probability distribution except when in the hot-hand or cold-hand state, there is no need to control for course-hole difficulty or player-quality.

6. I repeat the process 100 times and estimate the mean and the standard deviation of the coefficient estimate on the performance for the prior set of holes.
7. I then gauge what level of the hot hand would give the results observed in Table 2 for the hot- and cold-hand effects for nine-hole performance.

Note that these simulations are estimated under the presumption that the hot hand or cold hand lasts for a full tournament. This allows for a comparison of the magnitudes of the estimates from using various periods of performance, which should demonstrate how much measurement error contributes to reduced estimates with shorter periods of performance.

While a more realistic simulation would have players going in and out of the "cold" and "hot" states during the course of a round or tournament, the difficulty of modeling and interpreting such a simulation would likely make such an endeavor uninformative. The simple simulation developed here will demonstrate how much of a reduced or elevated performance would produce the true cold- and hot-hand estimates observed.

Figure 1 shows the differences in the probability of the possible outcomes on a single hole by various deciles of performance in a tournament. The top decile has a 27% chance of a birdie or better, compared to 22%, 20%, and 17% probability of a birdie or better for those in the 7^{th}, 4^{th}, and 1^{st} (bottom) deciles, respecitvely. And the probabilities of having a bogey or worse are 10% for the 10^{th} decile, 14% for the 7^{th} decile, 16% for the 4^{th} decile, and 21% for the 1^{st} decile.

Table 4 shows the implied score-relative-to-par for a round, given these probability distributions (which should match closely to the actual distribution of final tourna-

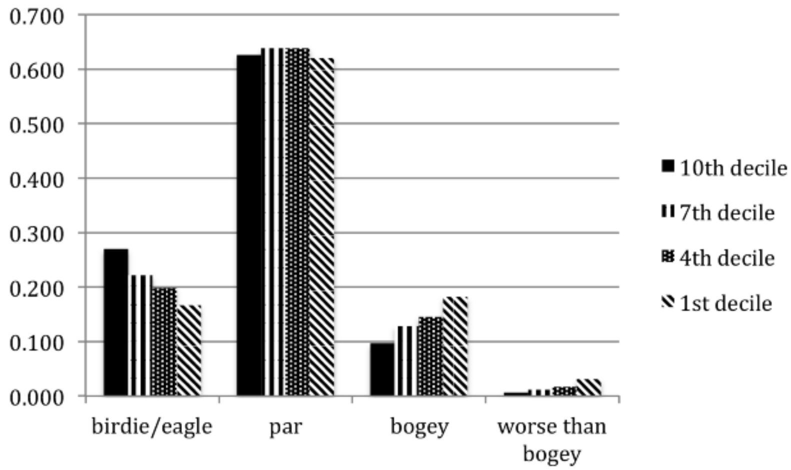

Figure 1. Probability of hole scores by selected finishing decile in a tournament.

Table 4. Implied Score Relative to Par, by Finishing Decile in a Tournament.

Finishing decile in a tournament	Implied score relative to par per round
10th	-3.02
9th	-2.16
8th	-1.74
7th	-1.32
6th	-1.01
5th	-0.70
4th	-0.38
3rd	-0.01
2nd	0.49
1st	1.41

ment scores). The implied scores are -3.0 for the 10th decile, -1.3 for the 7th decile, -0.4 for the 4th decile, and +1.4 for the 1st decile.

Using the Simulation to Gauge the Size of the Hot-Hand and Cold-Hand Effects
Table 5 shows the average coefficient estimates from different intensities of reduced performance (the cold hand) and elevated performance (the hot hand). I choose to evaluate the hot hand over nine holes, as it completes a full round of golf and it has the largest estimated cold-hand effect, and largest (insignificant) estimated hot-hand effect among the within-round models. The results would likely be fairly similar for examining 3-, 6-, or 18-hole performance.

The simulation results suggest that the cold hand can be quite forbidding. Under the scenario in which the cold hand occurs for 5% of all rounds of golf, the level of reduced performance that is most consistent with the cold-hand estimate of 0.043 in Table 2 is going from the 9th (2nd highest) decile to the 1st (lowest) decile. When the

Table 5. Cold- and Hot-Hand Scenarios Consistent with Estimated Effects for Nine-Hole Performance. (*n*=154.414)

	Cold hand (Model estimate = 0.043)		**Hot hand** (Model estimate = 0.007)	
Movement across deciles	5% frequency of col hand	10% frequency of cold hand	5% frequency of hot hand	10% frequency of hot hand
Reduced performance				
7th to 1st		0.043***		
9th to 1st	0.040***			
10th to 1st	0.061***			
Improved performance				
9th to 10th				0.004
8th to 10th			0.005**	0.010***
7th to 10th			0.010***	0.017***

Note: The cold-hand and hot-hand estimates are based on simple models as described in equation (2). The 10th decile is the best-performing, while the 1st decile is the worst-performing decile.
*** p<0.01, ** p<0.05, * p<0.1.

cold hand occurs 10% of the time, the 0.043 estimate is most consistent with going from the 7th to the 1st decile.

Such a cold-hand effect for nine-hole consecutive performance, of going from the 7th decile to the 1st decile, would involve, on each hole, a reduced probability of having a birdie or better of 5.4 percentage points, an increased probability of a bogey or worse of 7.3 percentage points, and a higher average score on 18 holes by 2.73 strokes per round.

For the hot-hand estimate from Table 2 of 0.007 (which is statistically insignificant), the level of elevated performance most consistent with that estimate, if the hot hand were to occur 5% of the time, is going from between the 7th and 8th decile to the 10th decile. If the hot hand were to occur 10% of the time, it would be going from between the 8th and 9th decile to the 10th decile. This is an important finding, as even with hundreds of thousands of observations, there may still be a real hot-hand effect that just cannot be detected due to the random nature of the outcome—which leads to misclassifications of the hot vs. normal state.

I should note here that these are just a few examples of hot- and cold-hand effects that would be consistent with what the various estimates would imply. And, of course, in any tournament, there would likely be players from all parts of the distribution (based on their abilities) having different levels of hot-hand or cold-hand performance (i.e., some would be playing well enough to move up or down one decile, while others may move up or down several deciles).

Table 6. Estimates on the Score-Relative-to-Par for the Prior Set of Holes, Using Simulated Data of Going from the 4th to 10th Decile for Tournament Performance

	5% occurrence of the hot hand	10% occurrence of the hot hand
18 holes		
score relative to par on prior 18 holes in the same tournament	0.041 (0.003)	0.075 (0.003)
9 holes		
score relative to par on prior 9 holes in the same round	0.021 (0.002)	0.039 (0.002)
6 holes		
score relative to par on prior 6 holes in the same round	0.014 (0.002)	0.026 (0.001)
3 holes		
score relative to par on prior 3 holes in the same round	0.007 (0.001)	0.013 (0.001)

Note: All estimates are statistically significant at the 1% level.

Using the Simulation to Demonstrate the Effects of Measurement Error

In Table 6, I show the estimated hot-hand effects from simulations going from the 4th decile of performance for tournament completers to the 10th (top-performing) decile.[1] This level of elevated performance is arbitrarily chosen. With a different level of elevated (or reduced) performance, the results would be different in levels, but the pattern would be the same from measuring performance over various sets of holes. The first set of results is based on the assumption that 5% of all tournaments are played in the hot-hand state. The second set of results is based on models where a player is hot in 10% of the tournaments.

Theoretically, with a constant hot-hand effect, it should be the same level of elevated performance from one set of holes to another, regardless of how many holes performance is measured over. But the results demonstrate that testing for the hot hand based on shorter periods of performance produces a smaller estimate due to randomness and luck playing a larger role in performance over shorter periods. Having half as many holes leads to an estimated effect that is just about one-half the size. Thus, although all of the estimates are statistically significant, it would generally be more difficult to detect significance for shorter periods of performance. This is entirely due to the randomness inherent in the outcomes, leading to measurement error for assignment of the hot and cold states. This has major implications for basketball hot-hand studies, which are subject to great measurement error due to limitations in the number of shots that can be used to categorize the player into the "hot" vs. normal state.

Conclusions

Golf can be an ideal sport to test for the hot hand given the standard set of holes played, the standard difficulty on a given set of holes for each player, and the relatively large set of possible outcomes allowing for separate tests for the hot hand and cold hand. At the same time, the concept of the hot hand in golf is different from that in other sports. In sports such as basketball, with the rush of adrenaline, the hot hand could occur from a player getting in the "zone," in which he/she reacts rather than thinks of the next move. Golf involves more concentration than other sports, as players have a considerable amount of time to think about the shot they are about to take and can even take some practice swings. Thus, any hot-hand effect would likely come from better concentration or perhaps greater confidence. Likewise a cold hand could be due to poor concentration or a loss of confidence.

Relative to prior studies on the golf hot hand, my contributions include the use of a fixed-effects model that allows all players to be included in one model; the examination of the hot- and cold-hand effects within a round of golf; and the use of actual scores rather than just dichotomous outcomes for being below par. Only one of the prior studies (Livingston, 2012) found any consistent evidence for a hot hand in golf, but only for a minor-league tour and not for the PGA tour. I also find no evidence for the hot hand on the PGA Tour, but I find evidence supporting the existence of a cold hand when examining performance over 3, 6, 9, and 18 holes.

The evidence for a cold hand but no evidence for a hot hand is quite plausible. Ben Hogan once said, "[Golf] is a game of misses. The guy who misses the best is going to win" ("Ben Hogan," para. 11). Having a hot hand would require a hot hand on all types of shots (drives, approaches, and putts), whereas a cold hand could occur with poor performance on just one type of shot, say putting.

One alternative explanation to these results is that a player having had performed poorly on prior holes may play riskier to make up for the lost ground. While this riskier play could pay off, on average it may lead to worse scores. For example, Stone and Arkes (2016) found evidence for riskier play (a greater probability of an above-par and below-par score) after an above-par score on the prior hole. This is more likely to be the case with just the very recent holes, which are fresher in a player's mind. But, it is always possible that players try to make up for weak performance on the prior six or nine holes. Nevertheless, the analysis here provides evidence consistent with a cold hand.

In light of the findings of Stone (2012) and Arkes (2013)—that measurement error due to the role of randomness causes downward biases on the estimates—I use simulations to gauge what level of elevated performance would create the (insignificant) estimate for a hot-hand effect and what level of reduced performance would create the (significant) estimates for the cold-hand effects. I find that a modest hot-hand effect could still exist and produce the insignificant estimate I find. But, the level of reduced performance that would produce the cold-hand effect I find is very large and frequent.

The lack of evidence for the hot hand does not indicate that there is no hot hand in golf. Rather, it merely indicates that there is no evidence for a hot hand from this study. But the evidence does suggest that if a hot hand in golf were to exist, it would be dwarfed by the size and frequency of the cold hand.

References

Albert, J., & Bennett, J. (2003). *Curve ball: Baseball, statistics, and the role of chance in the game.* New York, NY: Springer.

Albright, S. C. (1993). A statistical analysis of hitting streaks in baseball. *Journal of the American Statistical Association, 88*, 1175–1183.

Arkes, J. (2010). Revisiting the hot hand theory with free throw data in a multivariate framework. *Journal of Quantitative Analysis in Sports, 6*, 1–12.

Arkes, J. (2013). Misses in "hot hand" research. *Journal of Sports Economics, 14*, 401–410.

Baldson, E. M. (2013). Risk management with tournament incentives. *Journal of Quantitative Analysis in Sports, 9*, 301–317.

Ben Hogan. (n.d.). *Golf Today*. Retrieved from http://www.golftoday.co.uk/noticeboard/quotes/ben_hogan.html

Bocskocsky, A., Ezekowitz, J., & Stein, C. (2014). The hot hand: A new approach to an old "fallacy". Retrieved from http://www.sloansportsconference.com/wp-content/uploads/2014/02/2014_SSAC_The-Hot-Hand-A-New-Approach.pdf

Clark, R. D. (2003a). Streakiness among professional golfers: Fact or fiction? *International Journal of Sport Psychology, 34*, 63–79.

Clark, R. D. (2003b). An analysis of streaky performance on the LPGA tour. *Perceptual and Motor Skills, 97*, 365–370.

Clark, R. D. (2005). Examination of hole-to-hole streakiness on the PGA Tour. *Perceptual and Motor Skills, 100*, 806–814.

Connolly, R. A., & Rendleman, R. J., Jr. (2008). Skill, luck, and streaky play on the PGA Tour. *Journal of the American Statistical Association, 103*, 74–88.

Dorsey-Palmateer, R., & Smith, G. (2004). Bowlers' hot hands. *The American Statistician, 58*, 38–45.

Gibbons, C., & Suarez Serrato, J. C. (2011). Broken or fixed effects? Working paper.

Gilovich, T., Vallone, R., & Tversky, A. (1985). The hot hand in basketball: On the misperception of random sequences. *Cognitive Psychology, 17*, 295–314.

Hickman, D. & Metz, N. (2014). Performance under pressure on the PGA Tour. Working paper.

Livingston, J. A. (2012). The hot hand and the cold hand in professional golf. *Journal of Economic Behavior & Organization, 81*, 172–184.

Miller, J. B. & Sanjuro, A. (2014). A cold shower for the hot hand fallacy. Working paper. Retrieved from http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2450479

Rosenqvist, O., & Skans, O. N. (2014). Confidence enhanced performance—Evidence from professional golf tournaments. Working paper.

Smith, G. (2003). Horseshoe pitcher's hot hands. *Psychonomic Bulletin & Review, 10*, 753–758.

Stone, D. (2012). Measurement error and the hot hand. *The American Statistician*, *66*, 61–66.

Stone, D., & Arkes, J. (2016). Reference points, prospect theory and momentum on the PGA Tour. *Journal of Sports Economics*, forthcoming.

Vergin, R. C. (2000). Winning streaks in sports and the misperception of momentum. *Journal of Sport Behavior 23*, 181–197.

Endnote

[1] Testing for the cold hand rather than the hot hand would produce spiritually-similar results.